No tradeoff between versatility and robustness in gene circuit motifs

Payne, Joshua L

DOI: https://doi.org/10.1016/j.physa.2015.12.065

No tradeoff between versatility and robustness in gene circuit motifs

Joshua L. Payne[a,b,]

[a] *University of Zurich, Institute of Evolutionary Biology and Environmental Studies, Zurich, Switzerland*
[b] *Swiss Institute of Bioinformatics, Lausanne, Switzerland*

Abstract

Circuit motifs are small directed subgraphs that appear in real-world networks significantly more often than in randomized networks. In the Boolean model of gene circuits, most motifs are realized by multiple circuit genotypes. Each of a motif's constituent circuit genotypes may have one or more functions, which are embodied in the expression patterns the circuit forms in response to specific initial conditions. Recent enumeration of a space of nearly 17 million three-gene circuit genotypes revealed that all circuit motifs have more than one function, with the number of functions per motif ranging from 12 to nearly 30,000. This indicates that some motifs are more functionally versatile than others. However, the individual circuit genotypes that constitute each motif are less robust to mutation if they have many functions, hinting that functionally versatile motifs may be less robust to mutation than motifs with few functions. Here, I explore the relationship between versatility and robustness in circuit motifs, demonstrating that functionally versatile motifs are robust to mutation despite the inherent tradeoff between

Email address: joshua.payne@ieu.uzh.ch (Joshua L. Payne)

versatility and robustness at the level of an individual circuit genotype.

Keywords: Multifunctionality; regulatory circuits; signal-integration logic

1. Introduction

Gene regulatory networks are highly stylized, diagrammatic representations of the transcriptional and post-transcriptional mechanisms that cells use to control gene expression. In such networks, nodes represent genes and directed edges represent regulatory interactions between genes. One structural property that is common to the gene regulatory networks of organisms as different as yeast and human is the statistical enrichment of particular directed subgraphs known as circuit motifs (Lee et al., 2002; Alon, 2007; Boyle et al., 2014). Examples include three-gene motifs such as the feedforward loop and four-gene motifs such as the bi-fan (Milo et al., 2002) (Fig. 1). Experimental and theoretical analyses of these and other motifs have revealed their capacity to accelerate response times to intracellular signals (Mangan et al., 2006) and to buffer against transient fluctuations in gene expression levels (Prill et al., 2005), suggesting that the architecture of a circuit (*i.e.*, its motif) partly determines its function (Nykter et al., 2008b; Macía et al., 2009).

The function of a circuit is embodied in the expression pattern of its constituent genes — their level, timing, and location of expression. For example, a function of the gap gene circuit of *Drosophila melanogaster* is to form discrete bands of gene expression orthogonal to the anterior-posterior axis of the developing embryo, a function that is essential for the proper development of the fly's segmented body plan (Lawrence, 1992). Other examples of circuit

function include chemotaxis (Alon et al., 1999) and competence control (Süel et al., 2006) in bacteria, mating behavior in yeast (Tsong et al., 2003), lateral root development in plants (Chen et al., 2015), endoderm specification in the sea urchin (Hinman et al., 2003), and digit formation in the vertebrate limb (Raspopovic et al., 2014).

Gene circuits are often multifunctional, meaning that they form distinct expression patterns in different tissues or developmental stages, or in response to different combinations or levels of signaling molecules (Payne and Wagner, 2013). Said differently, multifunctional circuits drive multiple metastable expression states that are different from one another, and that are triggered by distinct sets of input signals. This phenomenon is exemplified by the *Hedgehog* gene circuit in butterflies, which both patterns the wing blade and helps to form the wing's eyespots (Keys et al., 1999). Other examples include the segment polarity network in *D. melanogaster*, which is involved in denticle patterning and the specification of neuroblasts (Carroll et al., 2001), and the circuit controlling mating behavior and the specification of cell type in yeast (Sorrells et al., 2015). Multifunctional circuits are also of interest to synthetic biologists, who engineer circuits to perform complex information processing tasks. For example, using transcription factors with engineered DNA binding domains, a circuit has been constructed that switches among the logical functions `AND` and `OR` in response to specific input signals (Gaber et al., 2014).

An important property of both natural and synthetic circuits is the robustness of their functions to genetic perturbation. Several theoretical and experimental studies have investigated the robustness of various gene cir-

cuits and networks (Little et al., 1999; Aldana and Cluzel, 2003; Ingolia, 2004; Voigt et al., 2005; Ma et al., 2006; Ciliberti et al., 2007b; Isalan et al., 2008; Rodrigo et al., 2011), yet we still know very little about the relationship between the architecture of a circuit and the robustness of its functions. This is mainly because earlier studies have focused on just one or a few circuit architectures, and only under a small subset of all possible initial conditions (Wall et al., 2005; Ingram et al., 2006; Conrad et al., 2008; Macía et al., 2009). Further, they did not consider multifunctional circuits, and they were limited to studying only a small fraction of the many regulatory programs that a given motif may implement. Such programs — referred to as signal-integration logic — are encoded in the regulatory regions of the circuit's genes, namely by the number, location, spacing, and orientation of transcription factor binding sites (Sharon et al., 2012; Smith et al., 2013), promoter strength (Lubliner et al., 2015), and other local sequence features (Raveh-Sadka et al., 2012; White et al., 2013; Levo et al., 2015). Mutations that alter a circuit's signal-integration logic may result in a new circuit function (Hunziker et al., 2010).

The *sin* operon in *Bacillus subtilis* provides an illustrative example of a circuit motif that can realize several distinct functions via changes in signal-integration logic (Voigt et al., 2005). The circuit's native function is a bistable switch that controls sporulation behavior, and the threshold of this switch can be fine-tuned via mutations in one of the circuit's two promoters. Mutations in the other promoter can lead to more drastic changes, transforming the circuit's function from a switch to a graded response, an oscillator, or a pulse generator. Importantly, these changes do not alter the circuit's architecture.

This motif is therefore highly versatile: small changes in signal-integration logic generate a diversity of circuit functions.[1]

It is not yet possible to experimentally characterize the functions of circuit motifs exhaustively (Schaerli and Isalan, 2013), so any comprehensive analysis of the relationship between circuit architecture and the robustness of circuit functions will necessitate the use of models. Kauffman's Boolean model (Kauffman, 1969) provides a useful framework for such an analysis. This is largely due to the model's explicit representation of a circuit's signal-integration logic, which determines the circuit's motif (Payne and Wagner, 2015) and its functions (Payne and Wagner, 2013). Moreover, for small circuits, it is possible to exhaustively enumerate all possible forms of signal-integration logic and under all possible initial conditions, facilitating the comprehensive exploration of the interplay between circuit architecture, circuit function, and the robustness of these functions to perturbation.

Previous work with the Boolean model has demonstrated a tradeoff between the number of functions (gene expression patterns) that an individual circuit may realize and the robustness of these functions to mutation (Payne and Wagner, 2013). Said differently, the more functions a circuit has, the less robust these functions are to genetic perturbation. Yet it remains to be seen whether this tradeoff also applies to circuit motifs, which are typically represented by many distinct circuits, each with their own signal-integration

[1]It is important to stress the difference between a circuit motif realizing multiple functions and an individual circuit being multifunctional. The former arises because motifs typically comprise many individual circuits, each with their own signal-integration logic, whereas the latter arises because individual circuits may yield different gene expression patterns in response to different initial conditions.

logic and functions (Payne and Wagner, 2015).

Here, I explore this potential tradeoff. To do so, I build upon my earlier work with three-gene Boolean circuits (Payne et al., 2014; Payne and Wagner, 2013, 2014, 2015), which has revealed five points that are relevant to the present study. First, there are nearly 17 million distinct forms of signal-integration logic, each of which I refer to as a circuit genotype. Second, many of these circuit genotypes have more than one function, *i.e.*, they are multifunctional. Third, as already mentioned, a tradeoff exists between the number of functions per circuit genotype and the robustness of those functions to mutation. Fourth, all circuit motifs have multiple functions, *i.e.*, collectively, the set of genotypes with a given motif always realize more than one function. And fifth, the number of functions per motif — *i.e.*, functional versatility — is highly variable, covering four orders of magnitude. Thus, circuit motifs vary in their versatility and comprise multifunctional circuit genotypes that individually exhibit a tradeoff between their number of functions and the robustness of these functions to mutation. The goal of this study is to determine whether the tradeoff between versatility and robustness also applies at the level of the circuit motif.

2. Model description

I consider fully-connected Boolean circuits with $N = 3$ genes (Fig. 2A). Circuits of this size are the typical focus of motif analyses (Milo et al., 2002, 2004) and drive important physiological and developmental processes, such as circadian oscillations in Cyanobacteria (Ishiura et al., 1998) and the specification of definitive hematopoiesis in the mouse embryo (Pimanda et al.,

2007). Despite its many simplifying assumptions, the Boolean model has provided important insight into a wide range of circuit dynamics, including the gene expression patterns of immune response in the macrophage (Nykter et al., 2008a) and the expression avalanches that result from gene knockouts in yeast (Serra et al., 2004, 2007).

Each gene in a Boolean circuit has its own signal-integration logic, which determines how its expression state will change in response to the 2^N possible combinations of expression states of the N genes in the circuit. In some cases, a gene's signal-integration logic may specify that its state is independent of the state of one or more genes in the circuit (a special case of what Kauffman et al. (2004) call "canalyzing rules"), thus rendering the corresponding regulatory interactions non-functional (gray arrows in Fig. 2A). For example, the signal-integration logic of a gene a may encode the statement "a `AND` b," which would render one of the three possible regulatory interactions ($c \rightarrow a$) non-functional. Similarly, the logical statement "not a" would render two regulatory interactions ($b \rightarrow a$ and $c \rightarrow a$) non-functional. Non-functional interactions can be pruned from the circuit without affecting circuit dynamics (Payne and Wagner, 2015). In this way, a circuit's signal integration logic encodes the circuit's motif.

Circuit dynamics are deterministic and arise from the synchronous updating of gene expression states, as prescribed by each gene's signal-integration logic (shown as look-up tables in Fig. 2A). The expression state of all genes in a circuit at time t is denoted as S_t. Starting from an initial state S_0, which represents the presence or absence of various signaling molecules or upstream regulatory factors, the circuit progresses through a series of states

until it reaches an equilibrium state S_∞ with period p, which may be a fixed-point ($p = 1$) or a cycle ($p > 1$).

Since a circuit's signal-integration logic fully specifies both circuit architecture and circuit dynamics, it is considered as the circuit's genotype G (Fig. 2B). Each circuit genotype is thus a binary vector of length $L = N \times 2^N = 24$ and the number of possible circuit genotypes is $2^L = 16,777,216$.

The function of a circuit is defined as a pairing of an initial and equilibrium state, $F = (S_0, S_\infty)$ (Fig. 2C). This definition is motivated by the functions of gene regulatory circuits in development and physiology, such as those that pattern the embryo in *Drosophila melanogaster* by interpreting a maternally-deposited morphogen gradient (S_0) to form discrete bands of gene expression along the developing embryo's anterior-posterior axis (S_∞) (Lawrence, 1992). Such functions are also relevant in robotics, where Boolean neural controllers interpret sensory information pertaining to the location of an object (S_0) to specify the coordinates of a robotic arm (S_∞) for grasping the object (Bongard, 2011)[2].

A circuit can have up to $k \leq 2^N$ functions, the set of which $\{F^{(1)} = (S_0^1, S_\infty^1), F^{(2)} = (S_0^2, S_\infty^2) \ldots F^{(k)} = (S_0^k, S_\infty^k)\}$ is referred to as a k-function or, if $k > 1$, as a multifunction (Payne and Wagner, 2013). I require that the equilibrium expression states of a k-function's k constituent functions are all fixed-point ($p = 1$) and different from one another ($S_\infty^1 \neq S_\infty^2 \neq \cdots \neq S_\infty^k$).

[2]While a definition of function that includes periodic equilibrium expression states ($p > 1$) is biologically sensible, especially for circuits controlling genetic oscillations, such as circadian rhythms (Young and Kay, 2001) and the cell cycle (Pomerening et al., 2005), it is my intention to investigate specific pairs of inputs (S_0) and outputs (S_∞) because this is typical of the circuits that inspire this model; those involved in development and physiology.

These requirements are motivated by developmental and physiological gene regulatory circuits, which typically specify fixed, temporally-invariant levels of gene expression in response to specific combinations of signaling molecules. For the three-gene circuits considered here, there are 32,399 possible k-functions, each of which is realized by at least one genotype (Payne and Wagner, 2013, 2014, 2015).

All possible three-gene circuit motifs are encoded by at least one of the nearly 17 million genotypes in this genotype space (Payne and Wagner, 2015). After correcting for graph isomorphisms, there are a total of 104 distinct motifs, including disconnected motifs. These motifs vary in the number of k-functions that are collectively realized by their constituent circuit genotypes (Payne and Wagner, 2015), and in the number of circuit genotypes they comprise. I refer to the latter as *motif abundance* — the number of genotypes per motif — a measure that is of central importance to this study. Other measures of interest include (i) the *functional repertoire* of a motif, defined as the set of unique k-functions realized by the circuit genotypes with the motif; (ii) the *versatility* of a motif, defined as the cardinality of the motif's functional repertoire; (iii) the *complexity* of a motif, defined as the number of regulatory interactions in the motif (Cotterell and Sharpe, 2010); and (iv) the *robustness* of a motif, defined as the average robustness of all of the k-functions that are realized by the motif's constituent genotypes.

For a given circuit genotype, the robustness of each of its k-functions is measured as the proportion of the genotype's mutational neighbors that have the same k-function (Payne and Wagner, 2013). This is determined by first removing the entries in the circuit's genotype G that correspond to

non-functional regulatory interactions, which yields a modified genotype G' (Fig. 2D). The k-functions of all mutational neighbors of G' — i.e., those genotypes that differ from G' at a single locus — are then assessed, and the fraction of these neighbors with the same k-function as G' is used as a measure of circuit robustness. Since most circuit genotypes have multiple k-functions, they also have multiple robustness values, one per k-function. The robustness of a motif is thus the average robustness of the motif's constituent genotypes, across each of the genotypes' k-functions (Fig. 2E).

3. Results

3.1. Motif abundance is highly variable

I first investigate motif abundance, the number of circuit genotypes per motif. Fig. 3A shows that motif abundance spans over 6 orders of magnitude. The two least abundant motifs are each realized by only 8 genotypes, and both are fully disconnected. One of these motifs has autoregulatory interactions on each gene, whereas the other does not have any regulatory interactions at all. The second least abundant motif is the feedback loop (Fig. 3a), which is realized by just 16 genotypes (Fig. 2A). The most abundant motif is fully connected (Fig. 3j), an architecture that is realized by over 10 million genotypes.

While there are 104 distinct circuit motifs, there are only 22 unique values of motif abundance (Fig. 3A). Thus, in many cases, different circuit motifs are equally abundant. Such motifs often share few structural similarities. For example, the feedforward loop (Fig. 3d) is as abundant as 21 other circuit motifs, but only 2 of these are simple variants of the feedforward design (*i.e.*,

they differ by the addition or deletion of a single regulatory interaction), despite the fact that all have nearly the same complexity (between 2 and 4 regulatory interactions).

Motif abundance is strongly correlated with motif complexity (Spearman's $r = 0.90, p = 1.61 \times 10^{-38}$; Fig. 3B, inset), indicating that more complex circuit motifs are generally represented by more circuit genotypes. Additionally, motifs with autoregulatory interactions are pervasive. In total, 85% of the 104 circuit motifs have at least one autoregulatory interaction and over 11 million circuit genotypes yield one of these motifs (71% of all possible genotypes).

3.2. Versatile motifs are abundant

I next investigate the relationship between motif versatility and motif abundance. Fig. 3B shows that as motif versatility increases, so does motif abundance (Spearman's $r = 0.80, p = 3.21 \times 10^{-24}$). This provides a simple and intuitive explanation for the previously observed variation in motif versatility (Payne and Wagner, 2015): some motifs comprise very few circuit genotypes, whereas others comprise very many. Motifs with more constituent genotypes have a greater diversity of signal-integration logic, which yields a larger functional repertoire.

3.3. Abundant motifs are robust

I next explore the relationship between motif robustness and motif abundance. Fig. 3C shows that these properties are positively correlated (Spearman's $r = 0.84, p = 3.87 \times 10^{-29}$), such that motif abundance increases exponentially with motif robustness. This might at first appear to be a

counterintuitive result, because the maximum number of functions $k_{\max}$ per k-function in a motif's functional repertoire (open circles in Fig. 3D) tends to increase with motif abundance (Spearman's $r = 0.47, p = 4.05 \times 10^{-7}$) and such k-functions are inherently less robust (Payne and Wagner, 2013). However, the average degree k_{avg} of the k-functions in a motif's functional repertoire (black line in Fig. 3D) is not correlated with motif abundance ($p = 0.09$). Thus, there are not enough k-functions of high k in the functional repertoires of complex motifs to bring down the average robustness of these functions.

3.4. Versatile motifs are robust

The preceding observations lead to the main result of this study: There is no tradeoff between versatility and robustness at the level of the circuit motif, despite the presence of the tradeoff at the level of the individual circuit genotype. In fact, versatility and robustness exhibit a synergistic relationship at the level of the circuit motif (Spearman's $\rho = 0.45, p = 1.86 \times 10^{-6}$), stemming from the positive correlation of these measures with motif abundance (Fig. 3B,C).

4. Discussion

I have used a Boolean model of gene regulatory circuits to investigate whether circuit motifs exhibit a tradeoff between the number of functions they realize — *i.e.*, their versatility — and the robustness of these functions to mutation. In contrast to individual circuit genotypes, which exhibit an inverse correlation between versatility and robustness (Payne and Wagner, 2013), no such tradeoff is observed for circuit motifs: The more functions a

circuit motif can realize, the more robust these functions are to mutation, on average. This synergistic relationship is mediated by motif abundance: Versatile motifs comprise a large number of circuit genotypes, and the functions of the genotypes of such abundant motifs tend to be robust to mutation.

The measure of motif abundance considered here is closely related to the concept of "designability" (Li et al., 1996), *i.e.*, the number of genotypes that yield a particular phenotype. For gene circuits, there are at least two aspects of designability (Rodrigo et al., 2011). First, there are multiple ways to construct the regulatory regions that yield a given circuit architecture (*i.e.*, motif), and second there are multiple forms of signal-integration logic that a circuit architecture may implement. The analysis presented here addressed the second aspect of designability, and suggests that complex motifs generally implement more forms of signal-integration logic than simple motifs. However, it is important to emphasize that complex circuit architectures are unlikely to be designable in natural systems. The reason is that they require complex regulatory architectures, which are less likely to evolve than the relatively simple regulatory architectures of simple motifs (Lynch, 2007). Nevertheless, such complex motifs have been identified, including the fully-connected three-gene circuit that drives hematopoietic development in the mouse embryo (Pimanda et al., 2007). For synthetic circuits, such designability constraints are relaxed, and the results presented here suggest that complex circuit motifs are better suited for the execution of complex information processing tasks than are simple motifs. Moreover, the functions of such circuits are likely to be robust to genetic perturbation.

Much recent work on gene regulatory circuits has focused on evolvability

(Aldana et al., 2007; Ciliberti et al., 2007a; Isalan et al., 2008; Greenbury et al., 2010; Garfield et al., 2013; Payne et al., 2014), defined as the ability to bring forth novel functions via small mutations in a circuit's coding and regulatory regions. These studies have shown that gene regulatory circuits tend to be highly evolvable, because mutations to the set of circuits with a given function give rise to a diversity of different circuit functions. Moreover, some these studies have uncovered a synergistic relationship between evolvability and robustness, similar to the relationship between versatility and robustness shown here. It is worth noting, however, that versatility and evolvability are orthogonal measures. While versatility also measures the number of functions that a given set of circuits may implement, this set of circuits is defined by their shared architecture, rather than by their shared function. Elucidating the relationship between versatility and evolvability presents an exciting direction for future research.

There are several caveats to this study. First, the Boolean model of regulatory circuits assumes that gene expression is binary. While this assumption sometimes provides a reasonable approximation of circuit functions (Mayo et al., 2006), it excludes the production and degradation rates of mRNA and protein, and thus precludes the study of important circuit functions such as "response accelerators" (Mangan et al., 2006) and "sign-sensitive delays" (Mangan and Alon, 2003). I was willing to accept this caveat because the assumption of binary gene expression facilitates a central goal of this study: To exhaustively characterize the versatility and robustness of all possible three-gene circuits, under all possible initial conditions and all possible forms of signal-integration logic. Second, the model assumes that gene expression

states are updated synchronously, which is clearly an oversimplification of the dynamics of biological circuits. Nonetheless, this assumption can be safely made because fixed-point equilibrium expression states are insensitive to the choice of synchronous *vs.* asynchronous updating scheme (Gershenson, 2002). Third, I did not consider periodic circuit functions, such as those involved in circadian rhythms and the cell cycle, partly because these functions *are* sensitive to the updating scheme (Gershenson, 2002). While including periodic circuit functions can only increase the versatility of circuit motifs, it may change the relationship between versatility and robustness because periodic functions are generally less robust than fixed-point functions (Payne and Wagner, 2013). Fourth, I did not consider environmental perturbation, such as gene expression noise, which is an important aspect of genetic regulation (Raser and O'Shea, 2005). The reason is that genetic and environmental robustness are positively correlated in the class of models studied here (Ciliberti et al., 2007b), and the former can therefore serve as a proxy for the latter.

Keeping these caveats in mind, the results presented here suggest that there is no tradeoff between versatility and robustness in gene circuit motifs, despite the presence of the tradeoff at the level of the individual circuit genotype.

5. Acknowledgments

I acknowledge support from the Ambizione program of the Swiss National Science Foundation. Thanks to Kathleen Sprouffske for comments on an earlier version of this manuscript.

References

Aldana, M., Balleza, E., Kauffman, S., Resendiz, O., 2007. Robustness and evolvability in genetic regulatory networks. J Theor Biol 245, 433–448.

Aldana, M., Cluzel, P., 2003. A natural class of robust networks. Proc Natl Acad Sci USA 100, 8710–8714.

Alon, U., 2007. Network motifs: theory and experimental approaches. Nat Rev Genet 8, 450–461.

Alon, U., Surette, M. G., Barkai, N., Leibler, S., 1999. Robustness in bacterial chemotaxis. Nature 397, 168–171.

Bongard, J., 2011. Spontaneous evolution of structural modularity in robot neural network controllers. In: Proceedings of the Genetic and Evolutionary Computation Conference. pp. 251–258.

Boyle, A. P., Araya, C. L., Brdlik, C., Cayting, P., Cheng, C. et al., 2014. Comparative analysis of regulatory information and circuits across distant species. Nature 512, 453–456.

Carroll, S. B., Grenier, J. K., Weatherbee, S. D., 2001. From DNA to Diversity. Molecular Genetics and the Evolution of Animal Design. Blackwell.

Chen, Q., Lui, Y., Maere, S., Lee, E., Van Isterdael et al., G., 2015. A coherent transcriptional feed-forward motif model for mediating auxin-sensitive *pin3* expression during lateral root development. Nat Comm 6, 8821.

Ciliberti, S., Martin, O. C., Wagner, A., 2007a. Innovation and robustness in complex regulatory gene networks. Proc Natl Acad Sci USA 104, 13591–13596.

Ciliberti, S., Martin, O. C., Wagner, A., 2007b. Robustness can evolve gradually in complex regulatory gene networks with varying topology. PLoS Computat Biol 3, e15.

Conrad, E., Mayo, A. E., Ninfa, A. J., Forger, D. B., 2008. Rate constants rather than biochemical mechanism determine behaviour of genetic clocks. J R Soc Interface 5, S9–S15.

Cotterell, J., Sharpe, J., 2010. An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. Mol Syst Biol 6, 425.

Gaber, R., Lebar, T., Majerle, A., Ster, B., Dobnikar, A., Bencina, M., Jerala, R., 2014. Designable DNA-binding domains enable construction of logic circuits in mammalian cells. Nat Chem Biol 10, 203–208.

Garfield, D. A., Runcie, D. E., Babbitt, C. C., Haygood, R., Nielsen, W. J., Wray, G. A., 2013. The impact of gene expression variation on the robustness and evolvability of a developmental gene regulatory network. PLoS Biol 11, e1001696.

Gershenson, C., 2002. Classification of random boolean networks. In: Eigth International Conference on Artificial Life. MIT Press, pp. 1–8.

Greenbury, S. F., Johnston, I. G., Smith, M. A., Doye, J. P. K., Louis, A. A.,

2010. The effect of scale-free topology on the robustness and evolvability of genetic regulatory networks. J Theor Biol 267, 48–61.

Hinman, V. F., Nguyen, A. T., Cameron, R. A., Davidson, E. H., 2003. Developmental gene regulatory network architecture across 500 million years of echinoderm evolution. Proc Natl Acad Sci USA 100, 13356–13361.

Hunziker, A., Tuboly, C., Horváth, P., Krishna, S., Semsey, S., 2010. Genetic flexibility of regulatory networks. Proc Natl Acad Sci USA 107, 12998–13003.

Ingolia, N. T., 2004. Topology and robustness in the *Drosophila* segment polarity network. PLoS Biol 2, 805–815.

Ingram, P. J., Stumpf, M. P. H., Stark, J., 2006. Network motifs: structure does not determine function. BMC Genomics 7, 108.

Isalan, M., Lemerle, C., Michalodimitrakis, K., Horn, C., Beltrao, P., Raineri, E., Garriga-Canut, M., Serrano, L., 2008. Evolvability and hierarchy in rewired bacterial gene networks. Nature 452, 840–846.

Ishiura, M., Kutsuna, S., Aoki, S., Iwasaki, H., Andersson, C. R., Tanabe, A., Golden, S. S., Johnson, C. H., Kondo, T., 1998. Expression of a gene cluster *kaiabc* as a circadian feedback process in Cyanobacteria. Science 281, 1519–1523.

Kauffman, S. A., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. J Theor Biol 22, 437–467.

399 Kauffman, S. A., Peterson, C., Samuelsson, B., Troein, C., 2004. Genetic
400 networks with canalyzing boolean rules are always stable. Proc Natl Acad
401 Sci USA 101, 17102–17107.

402 Keys, D. N., Lewis, D. L., Selegue, J. E., Pearson, B. J., Goodrich, L. V.,
403 Johnson, R. L., Gates, J., Scott, M. P., Carroll, S. B., 1999. Recruitment
404 of a *hedgehog* regulatory circuit in butterfly eyespot evolution. Science 283,
405 532–534.

406 Lawrence, P. A., 1992. The making of a fly. Blackwell.

407 Lee, T. I., Rinaldi, N. J., Robert, F., Odom, D. T., Bar-Joseph et al., Z., 2002.
408 Transcriptional regulatory networks in *saccharomyces cerevisiae*. Science
409 298, 799–804.

410 Levo, M., Zalckvar, E., Sharon, E., Machado, A. C. D., Kalma, Y., Lotam-
411 Pompan, M., Weinberger, A., Yakhini, Z., Rohs, R., Segal, E., 2015. Un-
412 raveling determinants of transcription factor binding outside the core bind-
413 ing site. Genome Res 25, 1018–1029.

414 Li, H., Helling, R., Tang, C., Wingreen, N., 1996. Emergence of preferred
415 structures in a simple model of protein folding. Science 273, 666–669.

416 Little, J. W., Shepley, D. P., Wert, D. W., 1999. Robustness of a gene regu-
417 latory circuit. The EMBO Journal 18, 4299–4307.

418 Lubliner, S., Regev, I., Lotan-Pompan, M., Edelheit, S., Weinberger, A.,
419 Segal, E., 2015. Core promoter sequence in yeast is a major determinant
420 of expression level. Genome Res 25, 1008–1017.

421 Lynch, M., 2007. The evolution of genetic networks by non-adaptive pro-
422 cesses. Nat Rev Genet 8, 803–813.

423 Ma, W., Lai, L., Ouyang, Q., Tang, C., 2006. Robustness and modular design
424 of the *Drosophila* segment polarity network. Mol Syst Biol, 70.

425 Macía, J., Widder, S., Solé, R., 2009. Specialized or flexible feed-forward loop
426 motifs: a question of topology. BMC Systems Biology 3, 84.

427 Mangan, S., Alon, U., 2003. Structure and function of the feed-forward loop
428 network motif. Proc Natl Acad Sci USA 100, 11980–11985.

429 Mangan, S., Itzkovitz, S., Zaslaver, A., Alon, U., 2006. The incoherent feed-
430 forward loop accelerates the response-time of the *gal* system of *Escherichia*
431 *coli*. J Mol Biol 356, 1073–1081.

432 Mayo, A. E., Setty, Y., Shavit, S., Zaslaver, A., Alon, U., 2006. Plasticity of
433 the *cis*-regulatory input function of a gene. PLoS Biol 4, e45.

434 Milo, R., Itzkovitz, S., Kashtan, N., Levitt, R., Shenn-Orr, S., Ayzenshtat,
435 I., , Sheffer, M., Alon, U., 2004. Superfamilies of evolved and designed
436 networks. Science 303, 1538–1542.

437 Milo, R., Shenn-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.,
438 2002. Network motifs: simple building blocks of complex networks. Science
439 298, 824–827.

440 Nykter, M., Price, N. D., Aldana, M., Ramsey, S. A., S A Kauffman et al.,
441 2008a. Gene expression dynamics in the macrophage exhibit criticality.
442 Proc Natl Acad Sci USA 105, 1897–1900.

Nykter, M., Price, N. D., Larjo, A., Aho, T., S A Kauffman et al., 2008b. Critical networks exhibit maximal information diversity in structure-dynamics relationships. Phys Rev Lett 100, 058702.

Payne, J., Moore, J., Wagner, A., 2014. Robustness, evolvability, and the logic of genetic regulation. Artif Life 20, 111–126.

Payne, J., Wagner, A., 2013. Constraint and contingency in multifunctional gene regulatory circuits. PLoS Comput Biol 9, e1003071.

Payne, J., Wagner, A., 2014. Latent phenotypes pervade gene regulatory circuits. BMC Syst Biol 8, 64.

Payne, J., Wagner, A., 2015. Function does not follow form in gene regulatory circuits. Sci Rep 5, 13015.

Pimanda, J. E., Ottersbach, K., Knezevic, K., Kinston, S., Chan et al., W. Y. I., 2007. Gata2, fli1, and scl form a recursively wired gene-regulatory circuit during early hematopoietic development. Proc Natl Acad Sci USA 104, 17692–17697.

Pomerening, J. R., Kim, S. Y., Ferrell, J. E., 2005. Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. Cell 122, 565–578.

Prill, R. J., Iglesias, P. A., Levchenko, A., 2005. Dynamic properties of network motifs contribute to biological network organization. PLoS Biology 3, e343.

Raser, J. M., O'Shea, E. K., 2005. Noise in gene expression: origins, consequences, and control. Science 309, 2010–2013.

Raspopovic, J., Marcon, L., Russo, L., Sharpe, J., 2014. Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. Science 345, 566–570.

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., Segal, E., 2012. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat Genet 44, 743–750.

Rodrigo, G., Carrera, J., Jaramillo, A., 2011. Computational design of synthetic regulatory networks from a genetic library to characterize the designability of dynamical behaviors. Nucleic Acids Res 39, e138.

Schaerli, Y., Isalan, M., 2013. Building synthetic gene circuits from combinatorial libraries: screening and selection strategies. Mol Biosyst 9, 1559–1567.

Serra, R., Villani, M., Graudenzi, A., Kauffman, S. A., 2007. Why a simple model of genetic regulatory networks describes the distribution of avalanches in gene expression data. J Theor Biol 246, 449–460.

Serra, R., Villani, M., Semeria, A., 2004. Genetic network models and statistical properties of gene expression data in knock-out experiments. J Theor Biol 227, 149–157.

Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., Segal, E., 2012. Inferring gene

regulatory logic from high-throughput measurements of thousands of systematically designed promoters. Nat Biotechnol 30, 521–530.

Smith, R. P., Taher, L., Patwardhan, R. P., Kim, M. J., Inoue, F., Shendure, J., Ovcharenko, I., Ahituv, N., 2013. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. Nat Genet 45, 1021–1028.

Sorrells, T. R., Booth, L. N., Tuch, B. B., Johnson, A. D., 2015. Intersecting transcription networks constrain gene regulatory network evolution. Nature 523, 361–365.

Süel, G. M., Garcia-Ojalvo, J., Liberman, L. M., Elowitz, M. B., 2006. An excitable gene regulatory circuit induces transient cellular differentiation. Nature 440, 545–550.

Tsong, A. E., Miller, M. G., Raisner, R. M., Johnson, A. D., 2003. Evolution of a combinatorial transcriptional circuit: A case study in yeasts. Cell 115, 389–399.

Voigt, C. A., Wolf, D. M., Arkin, A. P., 2005. The *bacillus subtilis sin* operon: An evolvable network motif. Genetics 169, 1187–1202.

Wall, M. E., Dunlop, M. J., Hlavacek, W. S., 2005. Multiple functions of a feed-forward-loop gene circuit. J Mol Biol 349, 501–514.

White, M. A., Myers, C. A., Corbo, J. C., Cohen, B. A., 2013. Massively parallel in vivo enhancer assay reveals that highly local features determine the *cis*-regulatory function of ChIP-seq peaks. Proc Natl Acad Sci USA 110, 11952–11957.

510 Young, M. W., Kay, S. A., 2001. Time zones: a comparative genetics of
511 circadian clocks. Nat Rev Genet 2, 702–715.

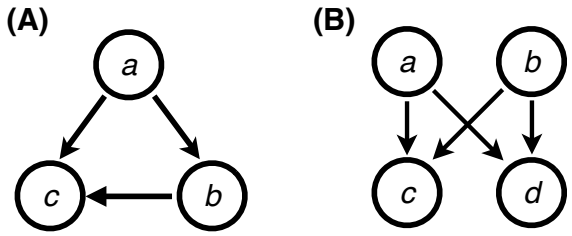

Figure 1: **Two examples of circuit motifs.** A gene regulatory circuit is a small subgraph of a larger gene regulatory network. Such circuits vary in their architecture (*i.e.*, the wiring diagram of "who" regulates "whom"). Each distinct architecture is referred to as a motif. For example, in the (A) feedforward motif, gene a regulates gene b and both genes a and b regulate gene c, whereas in the (B) bi-fan motif, genes a and b both regulate genes c and d.

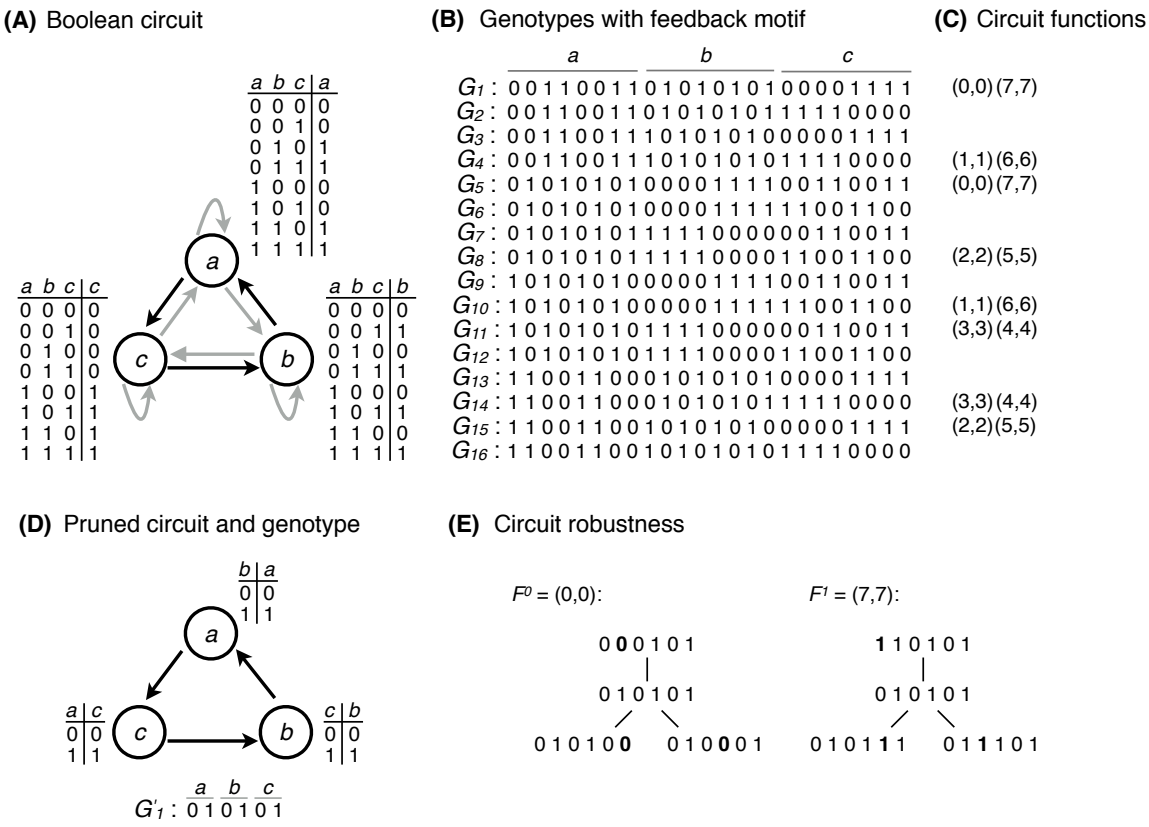

Figure 2: Caption on the following page.

Figure 2: **Schematic illustration of Boolean circuits.** (A) A Boolean circuit with $N = 3$ genes, labeled a, b, and c. Gene expression states are binary. The signal-integration logic of each gene is shown as a lookup table. These tables deterministically map the 2^N possible combinations of expression states of N genes to an output gene expression state. A directed edge $a \rightarrow c$ connects two genes if the expression state of c is dependent upon that of a (black arrows). Some signal-integration logic renders edges non-functional (gray arrows). For example, gene c simply mimics the output of gene a, regardless of its own state or that of gene b. The regulatory interactions $c \rightarrow c$ and $b \rightarrow c$ are therefore non-functional, as indicated by the gray arrows. By focusing on just the functional interactions (black arrows), it is evident that this circuit genotype encodes the feedback motif. (B) A circuit's genotype G is represented by a vector of length $L = N \times 2^N$. It is constructed by concatenating the rightmost columns of the lookup tables of the circuit's constituent genes. For example, the circuit shown in (A) has genotype G_1. There are 16 distinct circuit genotypes that encode the feedback motif, each with different signal-integration logic. (C) A circuit genotype may have between 0 and 2^N functions, each of which is a pairing of initial and equilibrium expression states. These states are shown here as integer representations of binary strings, *e.g.*, $(2, 1)$ represents the function $\langle 0, 1, 0\rangle \mapsto \langle 0, 0, 1\rangle$. For the feedback motif, eight genotypes $(G_2, G_3, G_6, G_7, G_9, G_{12}, G_{13}, G_{16})$ do not have any functions (*i.e.*, all of their equilibrium states have a period $p > 1$) and eight genotypes $(G_1, G_4, G_5, G_8, G_{10}, G_{11}, G_{14}, G_{15})$ have both monofunctions (*i.e.*, $k = 1$) and bifunctions (*i.e.*, $k = 2$), since a the functions of a circuit genotype can be expressed individually or in combination. The feedback motif's functional repertoire therefore comprises 8 monofunctions and 4 bifunctions; its functional versatility is 12. (D) Non-functional regulatory interactions can be pruned from the circuit shown in (A), yielding the modified genotype $G_1^{'}$. (E) The robustness of the circuit genotype is then assessed as the average robustness of its k-functions. This is calculated as the proportion of single-mutant neighbors that also have the k-function. In this example, the monofunctions $F^1 = (0, 0)$ and $F^2 = (7, 7)$ each have a robustness of 3/6, because $G_1^{'}$ has 6 mutational neighbors and three of them have the same monofunction. In contrast, the bifunction $\{F^1 = (0, 0), F^2 = (7, 7)\}$ has a robustness of 0 because none of its 6 neighbors have this bifunction. The robustness of the circuit genotype is therefore $(3/6 + 3/6 + 0)/3 = 1/3$.

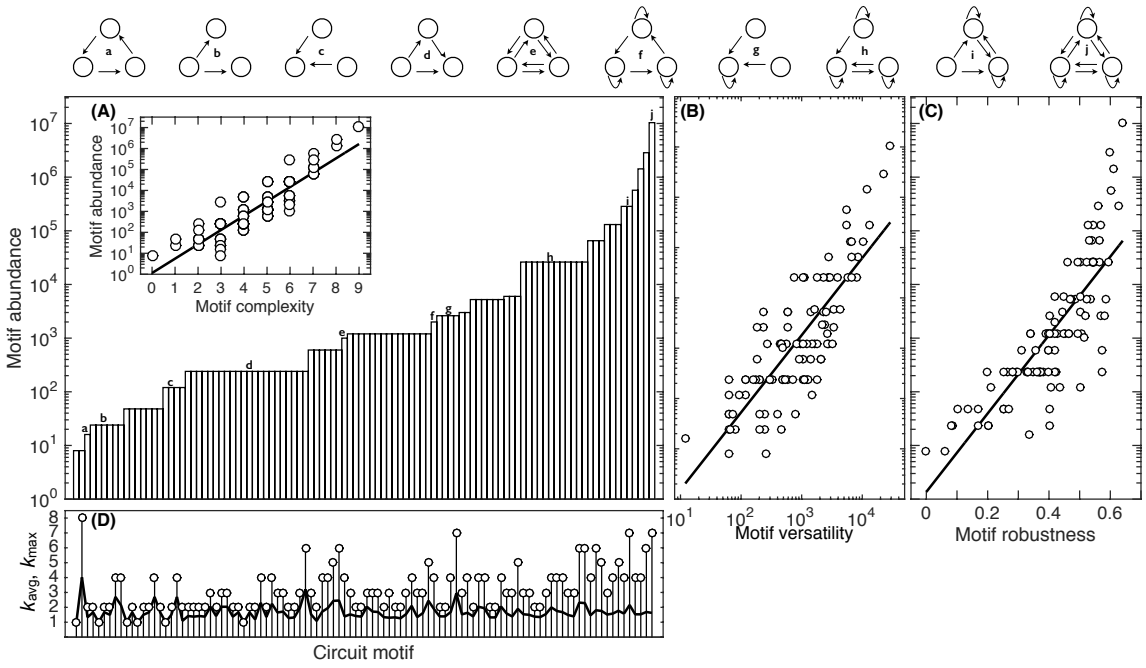

Figure 3: **Motif abundance, versatility, and robustness.** (A) Motif abundance per motif, arranged along the x-axis in increasing order of abundance. For reference, I draw 10 of these circuit motifs above the main panels, including the (a) feedback loop, (d) feedforward loop, and the (j) fully connected motif. The inset shows motif abundance in relation to motif complexity. (B) Motif abundance in relation to motif versatility for 104 three-gene motifs. (C) Motif abundance in relation to motif robustness for 104 three-gene motifs.. The y-axis in (B) and (C) is the same as in (A), and the solid lines in (A-C) represent the best fits to the data and are provided as visual guides. (D) The maximum ($k_{\max}$, open circles) and average (k_{avg}, solid line) degree of multifunctionality realized by the circuit genotypes that make up each of the 104 motifs, which are arranged along the x-axis as in (A).